STRUCTURAL GENOMICS OF *M. JANNASCHII*

Rapid Structural Determination of Proteins Using MAD Phasing

The goal of the Human Genome Project is to determine the DNA sequence of the human genome by the year 2005. One of the important objectives of determining the genomic sequences is to understand the cellular and molecular (biochemical and biophysical) functions of all the gene products (i.e., mostly proteins) encoded in the genomes, but the function of a protein cannot be readily inferred from the DNA sequence of a gene unless that sequence is significantly similar to that of a gene whose function is already known. The current estimate of the percentage of genes with gene products of known function varies from approximately 30% to 60%, depending on the genomes sequences. Furthermore, an even smaller fraction of the genes have gene products with known molecular functions. In structural genomics, researchers look for clues to the function of a protein in its three-dimensional structure.

Determining the structures of all the gene products of an organism would be an overwhelming task. Fortunately, the current database of protein structures strongly suggests that most proteins are classifiable in terms of a finite set of folds, the "folding basis set," and that each fold may be represented by a small number of biochemical or biophysical functions. Accordingly, large-scale projects to determined the structures of a few representatives from each fold family can provide a foundation for the functional genomics. This information can be combined with cellular functions derivable from mutational studies, transcription tracking, translation tracking, and interaction tracking.

To this end, researchers from UC Berkeley and Berkeley Lab are using the Macromolecular Crystallography Facility (MCF) at the ALS in a pilot study of the fully sequenced model hyperthermophilic archaebacterium *Methanococcus jannaschii*. The group has chosen several gene products from this organism—some with known cellular functions but without known molecular functions and some without any known functions—and have begun to determine their structures. The long-term goal of this project is to determine the structures of representative gene products in order to establish a folding basis set for the approximately 1800 gene products expressed in the microbe. The principal focus on finding a large number of new folds makes phase determination by multiple-wavelength anomalous diffraction (MAD) analysis a necessity.

Early results have already allowed the roles of two "hypothetical" proteins (proteins for which there is no other protein in the database with a gene having a similar sequence and a known function) to be tentatively identified from their structure alone. With data gathered at the MCF, for example, the group has determined the structure of hypothetical protein MJ0577 from *M. jannaschii*. The crystal structure was solved and refined within a few days after data collection was completed. The set of high-quality experimental phases from MAD measurements at the MCF has proven to be the key factor for interpreting and modeling the structures of the protein and ligands. For example, MJ0577 was identified as an ATP-binding protein after examination of the electron density map showed bound ATP.

The discovery of the ATP immediately narrows down the possible biochemical function of this protein. Biochemical experiments showed that MJ0577 has no appreciable ATPase activity by itself. However, when *M. jannaschii* cell extract was added to the reaction mixture, 50% of the ATP was hydrolyzed to ADP in 1 hour at 80˚C. This result indicates that MJ0577 requires one or more soluble components specific to *M. jannaschii* to stimulate ATP hydrolysis, suggesting that this is an ATP-mediated molecular switch analogous to Ras, a GTP-mediated molecular switch that requires GAP to hydrolyze GTP.

In these studies, the group has shown that MAD experiments can lead to very rapid protein-structure determination. Furthermore, in the case of MJ0577, the three-dimensional structure of unknown protein has provided direct information for functional (biochemical) assignment.

Sung-Hou Kim (510-486-4333), Physical BioSciences Division, E. O. Lawrence Berkeley National Laboratory and Department of Chemistry, University of California, Berkeley.

T. Zarembinski, L.-W. Hung, J. Mueller-Dieckmann, K.-K. Kim, H. Yokota, R. Kim, and S.-H. Kim, "Structure-based Assignment of the Biochemical Function of a Hypothetical Protein: a Test Case of Structural Genomics," *Proc. Natl. Acad.* Sci. 15 (1998) 15189

RESEARCH FUNDING: Office of Biological and Environmental Research, U. S. Department of Energy; Deutsche Forshungsgemeinshaft. Operation of the ALS is supported by the Division of Materials Sciences, U. S. Department of Energy.

STRUCTUAL GENOMICS OF *M. JANNASCHII*

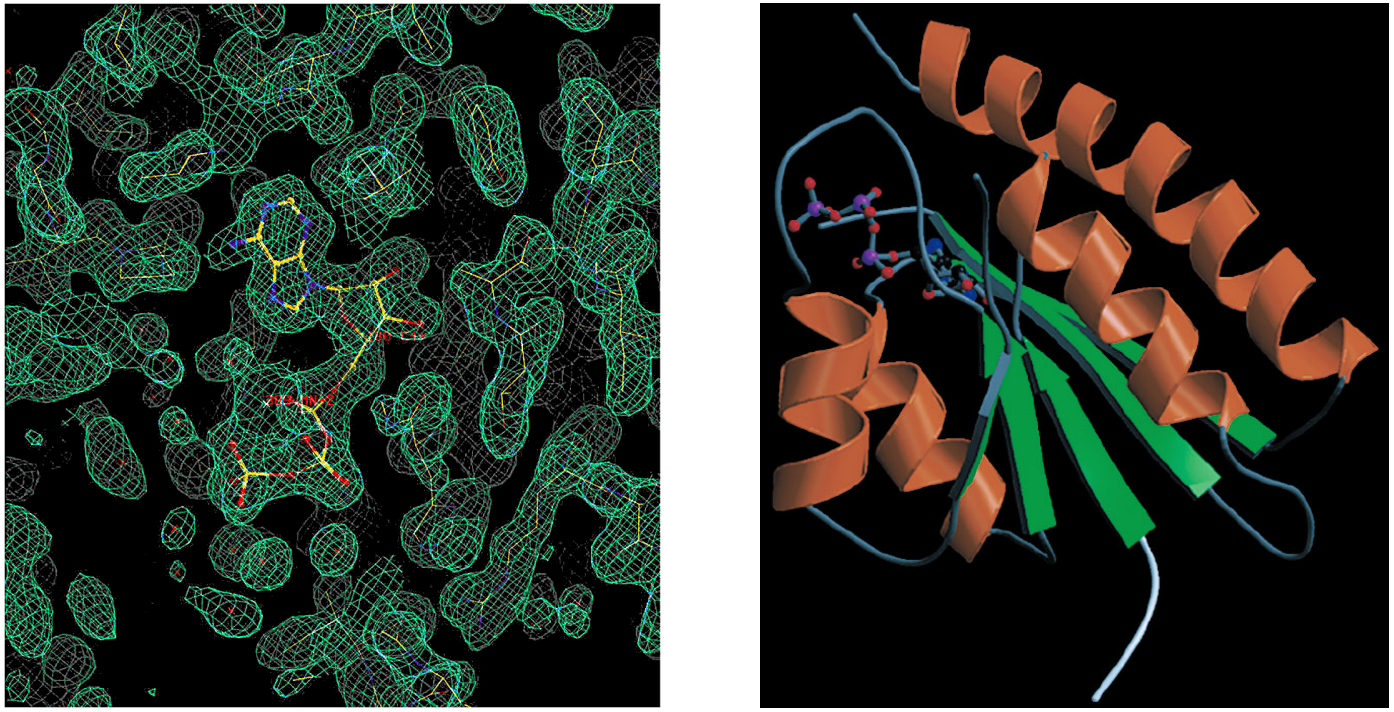
Rapid Structural Determination of Proteins using MAD Phasing

- **Human Genome Project goals**
  - *Determine the DNA sequence of the human genome*
  - *Understand the biological and biochemical functions of the gene products*
  - *Cannot always determine function from sequence alone*
- **Structural genomics**
  - *Determine protein function from its structure*
  - *Protein structures classifiable by a "folding basis set"*
  - *Correlate folding patterns with function of gene products*
- ***Methanococcus jannaschii***
  - *Fully sequenced thermophilic archaebacterium*
  - *Pilot project to test feasibility of structural genomics*
- **Two "hypothetical" proteins structures solved early in project**
  - *No proteins in gene database with a similar sequence and a known function*
  - *MAD experiments lead to rapid structure determination*
  - *Structure of MJ0577 provided direct information for functional assignment*

T. Zarembinski, L.-W. Hung, J. Mueller-Dieckmann, K.-K. Kim, H. Yokota, R. Kim, and S.-H. Kim (Berkeley Lab and University of California at Berkeley). [*Proc. Natl. Acad. Sci.* **15** (1998) 15189]

STRUCTUAL GENOMICS OF *M. JANNASCHII*

Rapid Structural Determination of Proteins using MAD Phasing



Structure of hypothetical protein MJ0577 solved at the Macromolecular Crystallography Facility. Left: Electron-density map derived from MAD experimental phases clearly shows a bound ATP. Right: The tertiary structure of MJ0577 is a nucleotide binding fold.

T. Zarembinski, L.-W. Hung, J. Mueller-Dieckmann, K.-K. Kim, H. Yokota, R. Kim, and S.-H. Kim (Berkeley Lab and University of California at Berkeley). [*Proc. Natl. Acad. Sci.* **15** (1998) 15189]